

Adaptive Filtering and Feature Detection Using Range Data

Clark F. Olson

Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Mail Stop 125-209, Pasadena, CA 91109-8099
<http://robotics.jpl.nasa.gov/people/olson/homepage.html>

Abstract

It is typical in edge and corner detection applications to examine a single scale or to consider some space of scales in the image without knowing which scale is appropriate for each location in the image. However, many images contain a wide variation in the distance to the scene points, and thus features of the same size can appear at greatly differing scales in the image. We present a method where the scale of the filtering and feature detection is varied locally according to the distance to the scene point, which we estimate through stereoscopy. The features that are detected are thus at the same scale in the world, rather than at the same scale in the image. This method has been implemented efficiently by filtering the image at a discrete set of scales and performing interpolation to estimate the response at the correct scale for each pixel. The application of this technique to an ordnance recognition problem has resulted in a considerable improvement in performance.

1 Introduction

Image smoothing and edge detection have been intensely studied subjects in computer vision and image processing. The selection of an appropriate scale for these processes is a problem that has received less attention. It is well known that using a single fixed scale over the entire image produces undesirable results, since edge phenomena occur at a multitude of scales. To alleviate this problem, techniques have been developed that examine the entire space of scales [1, 2, 3, 4] or that adaptively select a scale based on local image properties [5, 6, 7, 8]. However, the optimal method for combining the information from the scale-space is unclear. The scale selection methods that base their decisions on image properties, rather than the true scale at which the phenomena occur, can be confused by perspective effects.

In many applications it is desirable to detect image phenomena that are at the same scale in the world, which we call the *true scale*, rather than at the same scale in the image or by selecting a scale based on local image properties. Consider, for example, an image containing a textured surface in the foreground and an object of interest further from the camera. Techniques based on local image properties consider the textured surface at the scale it appears in the image. At this scale, the features might appear significant, even when this appearance is due only to perspective effects. If a method (such as stereoscopy) is available to determine the distance of the scene points from the camera, we can safely smooth these phenomena, while preserving the significant edges. Furthermore, if we seek objects of known size, the filtering and feature detection processes can be tuned to detect objects at the appropriate scale, regardless of their distance from the camera.

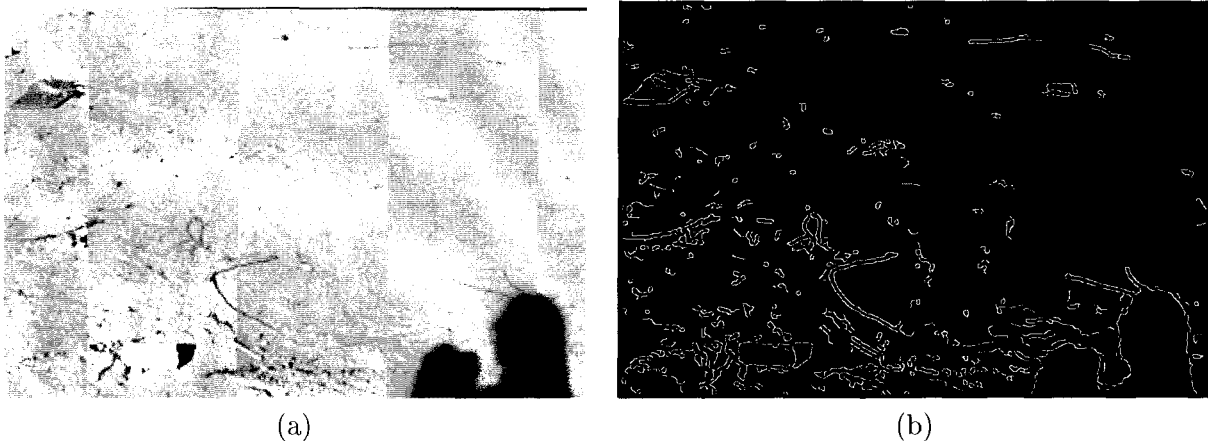


Figure 1: Motivating example. (a) Original image. (b) Edges detection after Gaussian smoothing with $\sigma = 2.0$.

A motivating example is shown in Figure 1. The image was collected at a live-fire test range near Nellis Air Force Base and contains two instances of live ordnance (one in the lower left, one in the upper right). In this case, the image was smoothed with a Gaussian filter with a constant scale of $\sigma = 2.0$ prior to differentiation and edge detection. It can be observed in the edge map, that the edges on the ordnance at close range are rough and there is considerable clutter in the foreground. However, when the scale of smoothing is increased, the edges of the ordnance in the background are no longer sharp. We wish to be able to smooth the image and perform edge detection such that the edges of both instances of ordnance are clear and there is little clutter in the edge map.

In addition to its value for scale selection, the range data is also useful for determining edge salience with respect to the scene characteristics. For example, edge salience measures such as length and straightness have been used [9]. However, the values these measures take are highly dependent on the distance of the edge from the camera. The stereo range information can be used to normalize these measures with respect to scene size and it is thus possible to determine edge salience with respect to the true scale rather than the image scale.

We have implemented these techniques using a variation of the Canny operator [10] to perform edge detection and a variation of the Förstner and Gülch operator [11] to perform feature detection. The filtering techniques are general and can be applied to most edge and feature detection methods. A mapping function between the distance to the pixel and the image scale is first determined. We next filter the image at a discrete set of scales. The response at each pixel at the appropriate scale is then interpolated from the discrete set of filter responses (similar to idea of steerable filters [12] or deformable kernels [13]). These responses are normalized, since the overall response to a general filter (e.g. a Gaussian derivative) is a function of the scale of the filter. Feature detection can then proceed according to the preferred operator.

2 Previous work

Since edges (and other features) appear at a wide variety of scales in an image, the concept of an image scale-space has been introduced [1, 4]. The scale-space can be defined as:

$$S(x, y, \sigma) = I(x, y) * g(x, y, \sigma) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(u, v) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-u)^2 + (y-v)^2}{2\sigma^2}} dudv.$$

While many researchers have noted the need to examine a variety of scales in the image, the means by which the scale-space is used is not a simple issue.

One method that has been investigated by Bergholm [14] is to track edges through the scale-space in a coarse-to-fine manner. The edges are detected at a coarse scale and progressively refined by the examination of smaller scales. Alternatively, Lu and Jain [3] have devised a complex system of rules for reasoning about edges in the scale-space, including six rules governing the progression of scales examined. When the scale of the edge is unknown, they recommend starting at a maximum scale σ_1 and decreasing the scale parameter by 1 pixel at each iteration.

Several methods of selecting a single local scale for each image pixel have been proposed. Jeong and Kim [6] select the local scales through the minimization of an energy functional over the scale-space using a regularization approach. The functional includes terms that encourage a large scale in uniform intensity areas, a small scale where intensities change significantly, and a smoothly varying scale over the image. Morrone *et al.* [8] suggest that the local scale should be a monotonically decreasing function of the gradient magnitude. They argue that this results in good localization through the use of a small scale when the contrast is high and good sensitivity using a large scale with the contrast is low. Lindeberg [7] notes that edge detection procedures seek to find maxima in the gradient magnitude in the spatial variables and that this principle can also be applied to the scale variable. He thus seeks the edge position in the scale-space where gradient magnitude is maximized.

Elder and Zucker [5] select the “minimum reliable scale”, which in their definition is the minimum scale at which the response level can be considered statistically reliable given the noise, edge amplitude, and image blur. This concept is used by Liang and Wang [15] to regulate an anisotropic diffusion process such that time at which diffusion ends is computed for each pixel according to the minimum reliable scale given by a local noise estimate. Marimont and Rubner [16] also use the minimum reliable scale, but according to a statistical confidence probability that they introduce.

Unlike these methods, we select the local scale of examination based on an estimate of the true scale, rather than trying to determine an appropriate scale through examination of the image. Our method is thus likely to yield better results when the real-world scale is the important one. As an alternative to selecting a single scale, our techniques can be used to complement scale-space techniques [4]. In this case, the range data would be used to transform the scale-space such that each scale plane was level with respect to the true scale rather than the image scale.

3 Depth acquisition

While any method that can associate range values with image pixels can be used with our adaptive filtering method, we concentrate on the use of stereoscopy to compute dense range maps of the

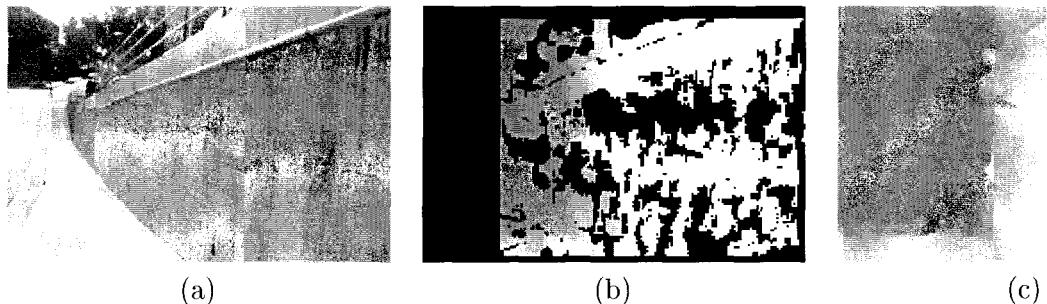


Figure 2: Range data extracted from a stereo pair. (a) Left image of a stereo pair. (b) Distance from the camera mapped into gray values. Black pixels indicate no valid range data. (c) Distances after filling pixels with no range data.

scene. The techniques that we use to compute the stereo range data have been described elsewhere [17, 18]. We briefly summarize this method here.

An off-line step, where the stereo camera rig is calibrated, is first performed. We use a camera model that allows arbitrary affine transformations of the image plane [19] and that has been extended to include radial lens distortion [20]. The remainder of the method is performed on-line.

At run-time, each image is first warped to remove the lens distortion and the images are rectified so that the corresponding scan-lines yield corresponding epipolar lines in the image. The disparity between the left and right images is measured for each pixel by minimizing the sum-of-squared-difference (SSD) measure of windows around the pixel in the Laplacian of the image. Subpixel disparity estimates are computed using parabolic interpolation on the SSD values neighboring the minimum. Outliers are removed through consistency checking and smoothing is performed over a 3×3 window to reduce noise. Finally, the coordinates of each pixel are computed using triangulation.

Note that not every pixel is assigned a range with this method. There are a number of factors that result in various pixels not being assigned a range, including occlusion, window effects, finite disparity limits, low texture, and outliers. Despite this problem, we must have a range estimate at each point in the image in order to estimate the scale that should be used for smoothing at that point. To resolve this dilemma, we propagate the range values from neighboring pixels using a simple method that approximates nearest neighbor search.

Figure 2 shows an example of the range data computed using these techniques. In this case, we fail to get range data at the left edge of the image, since this is the left image of a stereo pair, and there are significant areas over the rest of the image where the range data is discarded as not reliable. These values are filled with good estimates using the propagation techniques.

4 Filtering with variable scale

We perform variable-scale filtering using range data to select the appropriate scale at each pixel. The first step is to specify a mapping between the range data that has been computed for the scene and the scale at which the smoothing should be performed.

We map the range data into scales using:

$$\sigma(x, y) = \frac{K}{R(x, y)},$$

where $R(x, y)$ is the range computed at the image point (x, y) , $\sigma(x, y)$ is the scale to be used at (x, y) , and K is a pre-determined constant.

The constant, K , in this function can be determined using several methods. One possibility is to modify an automatic scale selection method (see, for example, [21]) to examine the image scale normalized by the depth values. A second possibility is to not limit ourselves to a single scale, but to consider the scale-space [4]. In this case, the scale-space can be warped such that the scale levels correspond to the true scale rather than the image scale. We use a third alternative. Since our primary application for these techniques is in detecting objects of known size, we select K based on the size of the objects.

In order to smooth the image prior to the application of the feature detection methods, we convolve the image with Gaussian filters. However, since we vary the scale at each pixel, the responses that we desire are governed by:

$$S(x, y) = \sum_{i=-W}^W \sum_{j=-W}^W I(x+i, y+j) \frac{1}{\sigma(x, y) \sqrt{2\pi}} e^{-\frac{i^2+j^2}{2\sigma(x, y)^2}},$$

where $I(x, y)$ is the image brightness at (x, y) and $2W + 1$ is the filter size.

Unfortunately, it is not efficient to compute this exactly for each image pixel. We perform this unconventional operation by convolving the image with a discrete set of Gaussian filters of various scales and interpolating the result at the appropriate scale for each pixel. This method for approximating a continuum of parameterized filters is similar to the techniques of steerable filters [12] and deformable kernels [13]. However, we have chosen parabolic interpolation rather than the linear combinations of the deformable kernels technique for simplicity and ease of implementation.

Since the range of scales that we are concerned with may be very large and Koenderink [1] has shown that a logarithmic sampling of the scale space is stable and in accordance with the principle that no scale should be preferred above others, we work in the $\log_2 \sigma$ domain. We have found that using discrete scales related by factors of two ($\sigma_n = 2^n \sigma_0$) is both convenient and effective.

The result of smoothing at each pixel with a filter of scale $\sigma(x, y)$ can be estimated through parabolic interpolation using the response of the discrete filter that is closest to the desired scale, $F_{\sigma_k}(x, y)$, and its two neighbors, $F_{\sigma_{k-1}}(x, y)$ and $F_{\sigma_{k+1}}(x, y)$. In determining an equation that yields the appropriate response, it is useful to perform a coordinate transform such that $z = \log_2 \frac{\sigma(x, y)}{\sigma_k}$. For $\sigma_{k-1} = \frac{1}{2}\sigma_k = \frac{1}{4}\sigma_{k+1}$, this yields $z_{k-1} = -1$, $z_k = 0$, and $z_{k+1} = 1$. With this transformation it is simple to show that the response we want is given by:

$$F(x, y) \approx az^2 + bz + c \tag{1}$$

$$a = \frac{1}{2}(F_{\sigma_{k-1}} - 2F_{\sigma_k} + F_{\sigma_{k+1}}) \tag{2}$$

$$b = \frac{1}{2}(F_{\sigma_{k+1}} - F_{\sigma_{k-1}}) \tag{3}$$

$$c = F_{\sigma_k} \tag{4}$$

$$z = \log_2 \frac{\sigma(x, y)}{\sigma_k} \tag{5}$$

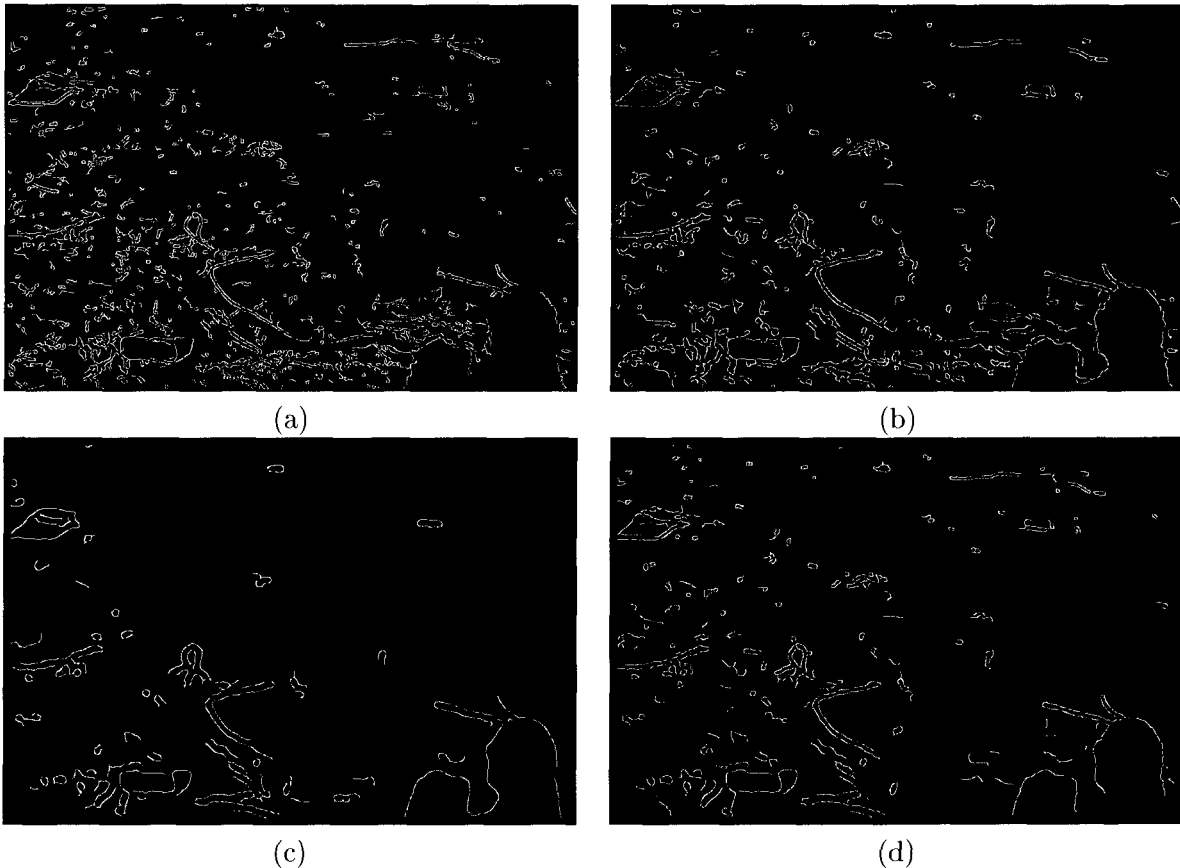


Figure 3: Edge detection results for the image in Figure 1. (a) Edges detected with $\sigma = 1.0$. (b) Edges detected with $\sigma = 2.0$. (c) Edges detected with $\sigma = 4.0$. (d) Edges detected with stereo-guided scale selection.

5 Edge detection

In our application of these techniques to edge detection, we follow the variable-scale smoothing described above with Canny’s edge detection method [10]. This technique computes the image gradients over the image in the x - and y -directions in order to determine the orientation and magnitude of the gradient at each pixel. Note, however, that if the gradient magnitudes are to be comparable, we must normalize them. This can be easily be seen by noticing that the response of a step edge to a Gaussian derivative filter varies with the scale of the filter. A Gaussian derivative aligned with a step edge yields a response proportional to $\frac{1}{\sigma}$. The gradient magnitudes will thus be stronger in the image regions that are smoothed at smaller scales if we do not normalize them. To correct this problem, we normalize the gradient magnitude at each pixel by multiplying by $\sigma(x, y)$.

Finally, non-maxima suppression is performed and the edges are detected using hysteresis thresholding. We determine the hysteresis thresholds adaptively through examination of the histogram of gradient magnitudes.

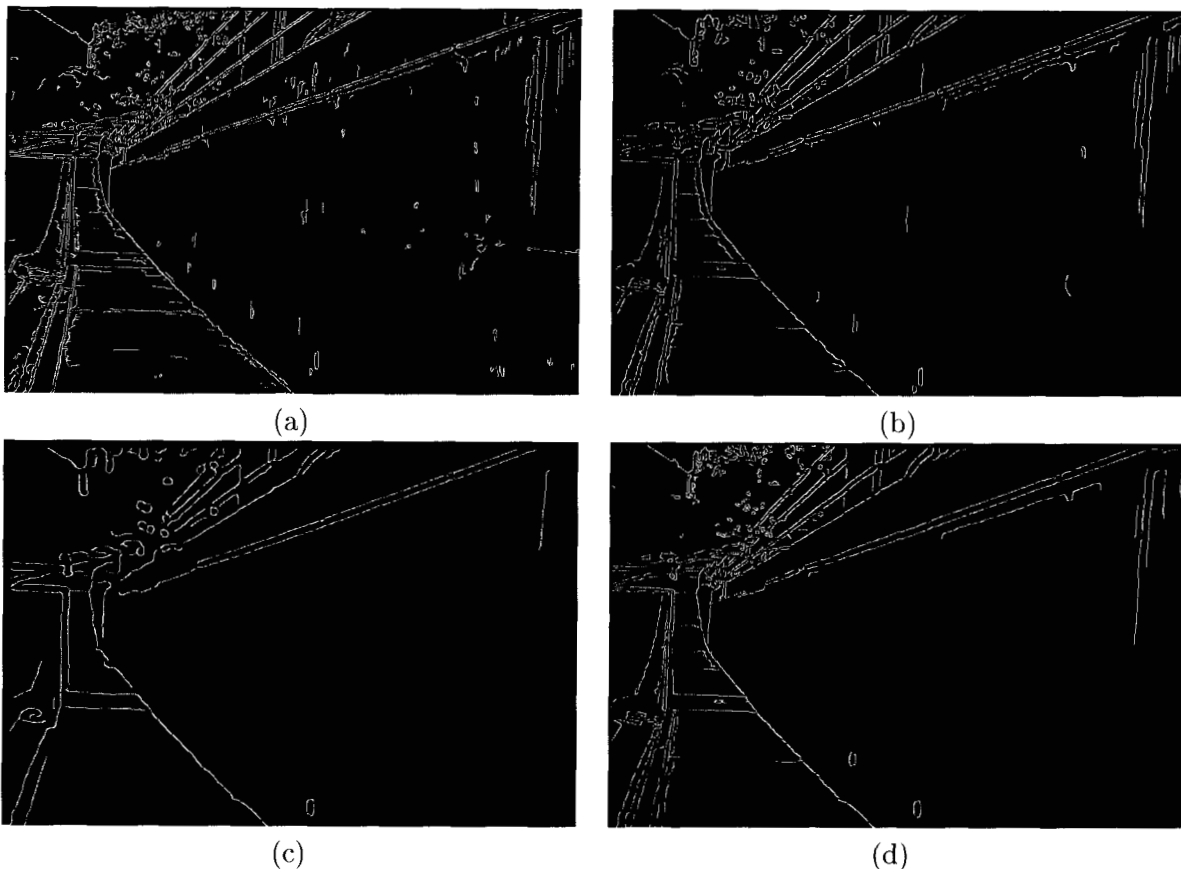


Figure 4: Edge detection results for the image in Figure 2. (a) Edges detected with $\sigma = 1.0$. (b) Edges detected with $\sigma = 2.0$. (c) Edges detected with $\sigma = 4.0$. (d) Edges detected with stereo-guided scale selection.

Figures 3 and 4 show an examples of edge detection with and without stereo-guided scale selection. The original images have 750×500 pixels and can be found in Figures 1 and 2. In these examples, the edges were detected at three scales ($\sigma = 1.0, 2.0, 4.0$) without the help of scale selection. Also given is the result with scale selection, where the response at each pixel was interpolated from the same three scales.

It can be seen that when a small scale ($\sigma = 1.0$) is used, many of the edges due to phenomena close to the camera are rough and a number of extraneous edges are detected due to the small scale, even though there is little image texture. However, when the scale is increased, we lose the details at the further phenomena (see, for example, the trees in the background and the end of the railing in Fig. 4). On the other hand, when the scale is selected adaptively using the stereo range map, we have good performance at both close and far edge phenomena. In this case, the edges that are detected are at the same scale in world, rather than the same scale in the image.

6 Edge salience evaluation

In addition to its use in performing edge detection, range data is also helpful in determining edge salience. Shorter edges that are detected at a larger distance are more likely to correspond to salient world edges than edges at close range that appear to be long due to perspective effects. We have primarily examined the summed gradient magnitude over the length of the edge and the local straightness of the edge as salience criteria, although many other salience measures could be used [9].

Consider, for example, a saliency measure where the gradient magnitude is summed along the length of the edge. The range data can be used to weight the gradient magnitude by the true edge length rather than the image edge length.¹ Alternatively, we could sum the ranges to the pixels (normalized appropriately for the field-of-view and edge direction) to estimate the length of the edge in the world coordinates.

As a second example, we may consider the local straightness of an edge at each of its edge pixels by examining the difference in the gradient direction at neighboring edge pixels along the edge. However, we would not expect identical edge phenomena appearing at different ranges to yield the same differences in the gradient direction between neighboring edges pixels. Edges closer to the camera will appear to be straighter, since the gradient differences will be smaller. To allow for this effect, the differences in gradient direction can be weighted by the range to the edge.

We have implemented both of these techniques, and they have resulted in a substantial improvement in our target application.

7 Feature detection

We have also applied these techniques to feature detection. In this case, we have applied the method of Förstner and Gülch [11] subsequent to the Gaussian smoothing. Since this operator also uses the image derivatives to determine whether a feature should be detected, we must normalize the gradients as was done for the edge detection. For each pixel, we then compute the circularity of the confidence ellipse q and the precision of the point w :

$$w = \frac{2 \cdot \det(N)}{\text{trace}(N)}$$

$$q = \frac{4 \cdot \det(N)}{\text{trace}(N)^2}$$

$$N = \begin{bmatrix} \left(\frac{\delta I}{\delta x}\right)^2 & \frac{\delta I}{\delta x} \cdot \frac{\delta I}{\delta y} \\ \frac{\delta I}{\delta x} \cdot \frac{\delta I}{\delta y} & \left(\frac{\delta I}{\delta y}\right)^2 \end{bmatrix}$$

The circularity q is between 0 and 1. We desire values closer to 1, since this indicates gradients in multiple directions, rather than a straight edge. A large w indicates the presence of strong gradients and measures the precision of the feature localization. We thus select as corners those positions that are local maxima and have w and q above some threshold value.

¹Note that, for non-frontal scenery, the orientation of the edge also affects the edge length. This effect can be accounted for if we estimate the three-dimensional orientation of the edge.

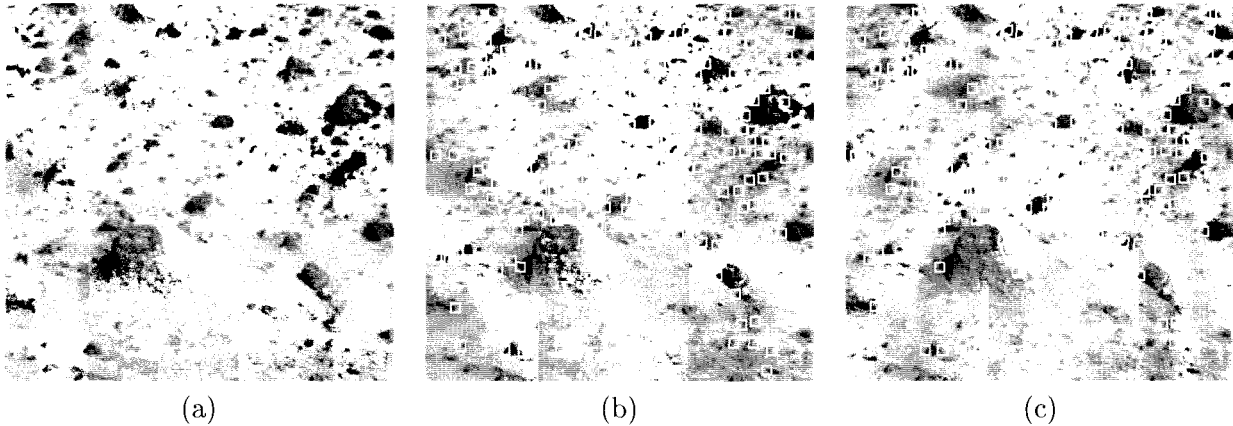


Figure 5: Feature extraction using adaptive scale selection. (a) Original image. (b) Corners detected with $\sigma = 2.0$. (c) Corners detected with adaptive scale selection.

Figure 5 shows an example of the application of these techniques to an image of Mars from the Pathfinder mission. It should be observed that when corners are detected with adaptive scale selection, less features are detected in the foreground, since they are less relevant. On the other hand, more features are detected further from the camera, where they are more relevant but appear smaller due to the perspective transformation.

8 Results

Our target application for these techniques is to recognize surface-lying ordnance in military test ranges using a stereo system mounted on an unmanned ground vehicle for the purpose of autonomous remediation. One method to evaluate the edge detection techniques is by the performance of this application when using the stereo-guided smoothing and edge detection versus the performance when it is not used. We have tested the techniques on a set of 48 gray-scale images consisting of barren terrain with an inert piece of ordnance present at various distances and orientations (see Figure 6).

In this experiment, we tested three scales individually ($\sigma = 0.8, 1.6, 3.2$), and the result with stereo-guided scale selection using the same three scales to interpolate from. After edge detection was performed, an algorithm to detect the ordnance using geometric cues was used to find candidate ordnance positions [22]. We also considered the combination of all of the candidates found at the three discrete scales (with duplicates removed).

Table 1 summarizes the results of this experiment. When the variable-scale smoothing and edge detection was performed, we achieved 40 correct recognitions out of the 48 cases. The eight failures occurred due to cases where the ordnance was at a significant distance from the camera and at an orientation nearly aligned with the camera axis. In addition, 18 false positives were detected in the images. Figure 6 shows two examples, one of which contains a false positive.

For each individual scale that was examined, we had more cases where the ordnance was missed than with stereo-guided scale selection and, in two of them, we also found more false positives.

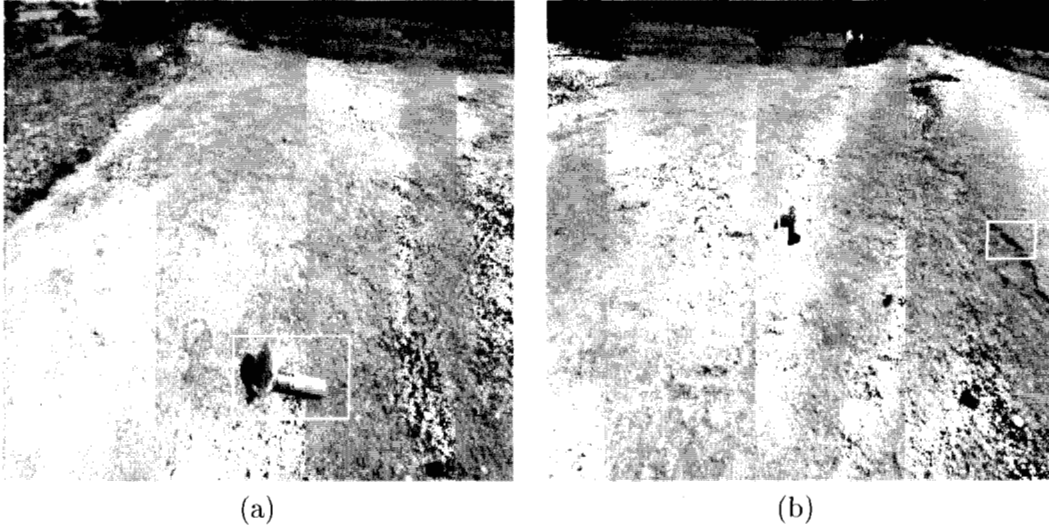


Figure 6: Ordnance recognition examples. (a) Correct detection at close range. (b) Correct detection at medium range and a false positive.

Scale	False negatives	False positives
1.0	12	28
2.0	11	28
4.0	13	14
all	7	45
variable	8	18

Table 1: Results in ordnance recognition application.

While $\sigma = 4.0$ found 4 less false positives, the detection performance was significantly degraded, since 5 additional ordnance instances were missed. When all of the candidates from the three scales were combined, there was one less false negative, but in this case the number of false positives rose sharply to 45.

Overall, the use of the stereo-guided scale selection techniques resulted in performance that was significantly superior to any of the individual scales or the combination of the scales.

9 Summary

We have described techniques that perform filtering and feature detection (edges and corners) adaptively using range data to select the scale at each pixel. This allows processing of the image to be performed with respect to the true scale of objects rather than the scale observed in the image. We have also used the range data for evaluating the salience of edges with respect to the true scale.

These techniques have been implemented by convolving the image with Gaussian derivatives at a discrete set of scales. These techniques have been implemented as a variation of the Canny edge

detector. The correct response at each image pixel is then estimated through parabolic interpolation of the known responses and normalization is performed so that the results are comparable across the image.

Improved results were obtained after the variable-scale filtering using both Canny's edge detector and the feature detection method of Förstner and Gülch. In addition, the application of this method to the problem of detecting unexploded ordnance in test ranges has resulted in a considerable improvement in performance.

Acknowledgments

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, and was sponsored by the Air Force Research Laboratory at Tyndall Air Force Base, Panama City, Florida, through an agreement with the National Aeronautics and Space Administration. Reference herein to any specific commercial product does not constitute or imply its endorsement by the United States Government, or the Jet Propulsion Laboratory, California Institute of Technology. A preliminary version of this work appeared in the 1998 IEEE Computer Society Conference on Computer Vision and Pattern Recognition [23].

References

- [1] J. J. Koenderink, "The structure of images", *Biological Cybernetics*, vol. 50, pp. 363–370, 1984.
- [2] T. Lindeberg, "Detecting salient blob-like image structures and their scales with a scale-space primal sketch: A method for focus-of-attention", *International Journal of Computer Vision*, vol. 11, no. 3, pp. 283–318, 1993.
- [3] Y. Lu and R. C. Jain, "Reasoning about edges in scale space", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 4, pp. 450–468, Apr. 1992.
- [4] A. P. Witkin, "Scale-space filtering", in *Proceedings of the International Joint Conference on Artificial Intelligence*, 1983, vol. 2, pp. 1019–1022.
- [5] J. H. Elder and S. W. Zucker, "Local scale control for edge detection and blur estimation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 7, pp. 699–716, July 1998.
- [6] H. Jeong and C. I. Kim, "Adaptive determination of filter scales for edge detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 5, pp. 579–585, May 1992.
- [7] T. Lindeberg, "Edge detection and ridge detection with automatic scale selection", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1996, pp. 465–470.
- [8] M. C. Morrone, A. Navangione, and D. Burr, "An adaptive approach to scale selection for line and edge detection", *Pattern Recognition Letters*, vol. 16, pp. 667–677, 1995.
- [9] P. L. Rosin, "Edges: Saliency measures and automatic thresholding", *Machine Vision and Applications*, vol. 9, pp. 139–159, 1997.
- [10] J. Canny, "A computational approach to edge detection", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 679–697, Nov. 1986.
- [11] W. Förstner and E. Gülch, "A fast operator for detection and precise locations of distinct points, corners, and centres of circular features", in *Proceedings of the Intercommission Conference on Fast Processing of Photogrammetric Data*, 1987, pp. 281–305.

- [12] W. T. Freeman and E. H. Adelson, "The design and use of steerable filters", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 9, pp. 891–906, Sept. 1991.
- [13] P. Perona, "Deformable kernels for early vision", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 5, pp. 488–499, May 1995.
- [14] F. Bergholm, "Edge focusing", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 6, pp. 726–741, Nov. 1987.
- [15] P. Liang and Y. F. Wang, "Local scale controlled anisotropic diffusion with local noise estimate for image smoothing and edge detection", in *Proceedings of the International Conference on Computer Vision*, 1998, pp. 193–200.
- [16] D. H. Marimont and Y. Rubner, "A probabilistic framework for edge detection and scale selection", in *Proceedings of the International Conference on Computer Vision*, 1998, pp. 207–214.
- [17] L. Matthies, "Stereo vision for planetary rovers: Stochastic modeling to near real-time implementation", *International Journal of Computer Vision*, vol. 8, no. 1, pp. 71–91, July 1992.
- [18] L. Matthies, A. Kelly, T. Litwin, and G. Tharp, "Obstacle detection for unmanned ground vehicles: A progress report", in *Proceedings of the International Symposium on Robotics Research*, 1996, pp. 475–486.
- [19] Y. Yakimovsky and R. Cunningham, "A system for extracting three-dimensional measurements from a stereo pair of TV cameras", *Computer Vision, Graphics, and Image Processing*, vol. 7, pp. 195–210, 1978.
- [20] D. B. Gennery, "Least-squares camera calibration including lens distortion and automatic editing of calibration points", in *Calibration and Orientation of Cameras in Computer Vision*, A. Grün and T. S. Huang, Eds. Springer-Verlag, in press.
- [21] E. R. Hancock and J. Kittler, "Adaptive estimation of hysteresis thresholds", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1991, pp. 196–201.
- [22] C. F. Olson and L. H. Matthies, "Visual ordnance recognition for clearing test ranges", in *Detection and Remediation Technologies for Mines and Mine-Like Targets III, Proc. SPIE*, 1998.
- [23] C. F. Olson, "Variable-scale smoothing and edge detection guided by stereoscopy", in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1998, pp. 80–85.